

## **TITLE:**

### **Determining a Compact Model to Transcribe the Arabic Language Acoustically in a Well Defined Basic Phonetic Study**

#### Field of Invention

The present invention relates generally to the field of controlling a computer dictation application using multi-gender human voice instead of a keyboard. More specifically, the present invention is related to determining a compact model to transcribe the Arabic language acoustically in a well-defined basic phonetic study.

#### Background of the Invention

Phonetics, as defined by the Merriam-Webster® dictionary (Collegiate 10th ed.), is a system of speech sounds of a language or group of languages, and further comprises the study and systematic classification of the sounds made in spoken utterance. Hence, the phonetic system represents the practical application of this science to language study. An important part of phonetics is phonemes.

Phonemes, as defined by Merriam-Webster® dictionary (Collegiate 10th ed.), are abstract units of the phonetics system (associated with a particular language) that correspond to a group of speech sounds. For example, velar  $|k|$  of *cool* and the palatal  $|k|$  of *keel* are distinct sounds in the English language and are part of a set similar speech sounds. Another term related to Phonemes is allophones.

Allophone, as defined by Merriam-Webster® dictionary (Collegiate 10<sup>th</sup> ed.), is one of two or more variants of the same phoneme. For example, the aspirated  $|p|$  of pin and the unaspirated  $|p|$  of spin are allophones of the phoneme  $|p|$ .

Orthography is another system associated with the sounds of a given language. Orthography, as defined by Merriam-Webster® dictionary (Collegiate 10<sup>th</sup> ed.), is the representation of the sounds of a language by letters and diacritics. A diacritic is further defined as a mark near or through an orthographic or phonetic character or combination of characters indicating a phonetic value different from that given the unmarked or otherwise marked element. An example of a diacritic is the acute accents of *résumé*, which are added to the letter *e* to indicate a special phonetic value.

Additionally, some foreign languages often use diacritics to double the force of the phoneme, and they further use geminated graphemes. Graphemes are the set of units of a writing system (as letters and letter combinations) that represent a phoneme. Geminated graphemes are a sequence of identical speech sounds (as in *meanness* or Italian *notte*).

One of the advances in recent years is the impact of computers in the field of phonetics. One of the major challenges associated with human speech and computers is automatic speech recognition or ASR. ASR is defined as the ability of a computer-based system to recognize and decipher human voice. ASR systems are usually programmed to recognize a simple set of words that are common to a group of users, or sometimes ASR systems are programmed to recognize a complex set of words associated with a specific user.

One common problem associated with phonetic representation of foreign language (such as Arabic) is the abundance of phonetics associated with such languages. Arabic language displays a difference in orthography and phonetics

associated with the language. This is best illustrated by the example of geminated graphemes. The feature of gemination is inherent in most Arabic phonetic alphabets, which is defined as doubling the force of the phoneme that is marked by a superscript sign. During editing, people do not write this sign unless it is crucially needed to decipher a certain meaning from another. That's why the grapheme is written only once. Another example is that the language exhibits different variations between the kinds of vowel distribution; either they are short or long vowels. Tables 2 and 3, as detailed hereafter, represent the different features of vowels and gemination consecutively. Thus, software representing such a system comprising a myriad of phonetics inevitably requires a significant allocation of memory on a computer-based device for storage of such plurality of phonetics.

A variety of software applications are available today that utilize the phonetics system to recognize the speech of human users. But, none of the prior art software utilizes an automatic speech recognition system that uses an orthographic system comprising a compact set of phonetics. Whatever the precise merits, features and advantages of the above cited references, none of them achieves or fulfills the purposes of the present invention.

#### SUMMARY OF THE INVENTION

The present invention provides for a method and a system for developing a compact model to transcribe the Arabic language acoustically based on a well-defined basic phonetic study. The compact model is accomplished in the present invention by reducing the set of phonemes. Thus, the creation of a minimized set of phonemes

helps in reducing memory consumption, hence a fast execution of word editing. Table 4 represents the minimized set used in the dictation system. More specifically, Arabic words, provided as examples in Tables 1 and 4, illustrate that in the instance of gemination, only one grapheme (and not a doubled one) is used, while it is still doubled phonemically. It is also clear in the case of vowels; that is, while there are almost six degrees of vowels in table 1, and in table 4 there are only three. Hence, the difference in pronunciation is not taken into account in the written text. Accordingly, the present invention provides for a set of phonemes to be used by Arabic dictation software capable of automatic speech recognition.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the method associated with the preferred embodiment of the present invention for determining a compact model to transcribe the Arabic language acoustically based on a well-defined basic phonetic study.

Figure 2 illustrates in further detail the data extraction step of Figure 1.

Figure 3 illustrates the composition of the maximal set described in the method of Figure 1.

Figure 4 illustrates the various kinds of phonemes.

#### DESCRIPTION OF THE PREFERRED EMBODIMENTS

While this invention is illustrated and described in a preferred embodiment, the invention may be produced in many different configurations, forms and materials. There is depicted in the drawings, and will herein be described in detail, a preferred embodiment of the invention, with the understanding that the present disclosure is to

be considered as an exemplification of the principles of the invention and the associated functional specifications for its construction and is not intended to limit the invention to the embodiment illustrated. Those skilled in the art will envision many other possible variations within the scope of the present invention.

One important initial step involved in the development of automatic speech recognition (ASR) software is a "basic phonetic study". A general description of such a study starts with identifying a language on which a basic phonetic study needs to be performed, any material related to the phonology and phonetics of the identified language is collected (or alternatively extracted from a database over a network). This provides for an overview of the phonetic structure of the identified language. Furthermore, technological problems and transcription problems associated with the language are identified. For example, literature in Arabic phonetics uses the terms "emphatic", "pharyngealized", and "velarized", which exhibit clear differences that mark their uniqueness. Additionally, it is necessary to interpret the symbols in the literature and find a mapping to a single and more recent phonetic alphabet based on feature description rather than symbol shapes.

It should be noted that the International Phonetic Alphabet (IPA) was used in conjunction with this invention. The IPA, as defined by the International Phonetic Association (<http://www.arts.gla.ac.uk/IPA/ipa.html>) is a standard set of symbols for transcribing the sounds of spoken languages. The above mentioned website provides for a full chart of IPA symbols as reproduced below. Furthermore, charts for

consonants, vowels, tones and accents, suprasegmentals, diacritics and other symbols are also provided. The last version of the IPA dates to 1993, as shown below:

# THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

## CONSONANTS (PULMONIC)

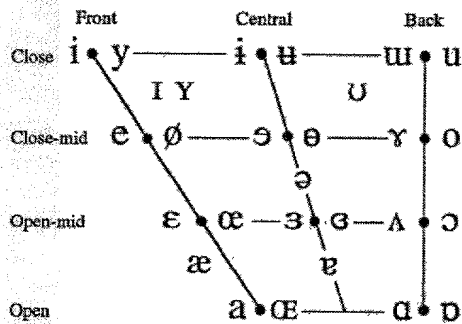
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

## CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
⦿ Bilabial	ɓ Bilabial	as in:
ǀ Dental	ɗ Dental/alveolar	p' Bilabial
ǃ (Post)alveolar	ɟ Palatal	t' Dental/alveolar
ǂ Palatoalveolar	ɡ Velar	k' Velar
ǁ Alveolar lateral	ɠ Uvular	s' Alveolar fricative

## VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

## OTHER SYMBOLS

ɱ Voiceless labial-velar fricative	ɕ ʑ Alveolo-palatal fricatives
ɰ Voiced labial-velar approximant	ɺ Alveolar lateral flap
ɰ Voiced labial-palatal approximant	ɺ Simultaneous ʃ and x
ħ Voiceless epiglottal fricative	Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʕ Voiced epiglottal fricative	
ʕ Epiglottal plosive	

kp ts

## SUPRASEGMENTALS

	Primary stress	Secondary stress	Long	Half-long	Extra-short	Syllable break	Minor (foot) group	Major (intonation) group	Linking (absence of a break)
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ
	ˈ	ˌ	ː	ˑ	˚	ˌ.ækt	ˌ	ˌ	ˌ

## TONES & WORD ACCENTS

LEVEL	CONTOUR
˥ Extra high	˩ Rising
˨ High	˩ Falling
˨ Mid	˩ High rising
˨ Low	˩ Low rising
˨ Extra low	˩ Rising-falling etc.
˩ Downstep	˩ Global rise
˩ Upstep	˩ Global fall

## DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. ɲ̥

˚ Voiceless	̤ ̥	˚ Breathy voiced	̤ ̥	˚ Dental	̤ ̥
˚ Voiced	̤ ̥	˚ Creaky voiced	̤ ̥	˚ Apical	̤ ̥
˚ Aspirated	̤ ̥	˚ Linguolabial	̤ ̥	˚ Laminal	̤ ̥
˚ More rounded	̤ ̥	˚ Labialized	̤ ̥	˚ Nasalized	̤ ̥
˚ Less rounded	̤ ̥	˚ Palatalized	̤ ̥	˚ Nasal release	̤ ̥
˚ Advanced	̤ ̥	˚ Velarized	̤ ̥	˚ Lateral release	̤ ̥
˚ Retracted	̤ ̥	˚ Pharyngealized	̤ ̥	˚ No audible release	̤ ̥
˚ Centralized	̤ ̥	˚ Velarized or pharyngealized	̤ ̥		
˚ Mid-centralized	̤ ̥	˚ Raised	̤ ̥ (ɹ̥ = voiced alveolar fricative)		
˚ Syllabic	̤ ̥	˚ Lowered	̤ ̥ (ɹ̥ = voiced bilabial approximant)		
˚ Non-syllabic	̤ ̥	˚ Advanced Tongue Root	̤ ̥		
˚ Rhoticity	̤ ̥	˚ Retracted Tongue Root	̤ ̥		

In the present invention, all units regarding the literature of the language in question are collected (or alternatively extracted via a database). Next, all unwanted elements are removed. This compilation allows one to establish feature sets required to describe each and all sounds of the language, and describe accurately each phonological or phonetic unit associated with the language. After the feature set and unit transcription, a representational symbol of the transcription alphabet is selected.

Subsequently, a structured table is constructed with the following information: i) all phonemes of the language, ii) all allophones of the language and their relation to the phonemes, iii) a preliminary set of rules governing the selection of allophones, iv) a set of examples, and v) the most common representation of the sounds using Roman letters.

Figure 1 illustrates the method **100** associated with the preferred embodiment of the present invention for determining a compact model to transcribe the Arabic language acoustically (based on a well-defined basic phonetic study). First, a language for which a compact model is to be developed is identified **102**. Next, information regarding the identified language is extracted or collected **104**. Data extraction can be accomplished via a variety means including, but not limited to: extracting data regarding the Arabic language via a network (such as the Internet, Local Area Network (LAN), Wide Area Network (WAN) or database (local or remote). Next, from the extracted data, a list is created where the phonological and phonetic units are defined **106**. As a next step, the variations in the Arabic language are identified **108**. For example, variations in classical Arabic, Modern Standard Arabic (MSA), and colloquial



Arabic are identified. Next, a maximal set is created that contains all phonemes, allophones, and transliteration symbols associated with the Arabic language **110**. Transliteration refers to the process of representing or spelling a word (in a first language) in the characters of another alphabet (second language). Lastly, the maximal set is reduced **112** to provide for a compact set to transcribe the Arabic language acoustically. The details of the reduction step are explained in detail in the following sections.

The data extraction step of Figure 1 (**102**) is illustrated in further detail in Figure 2. With the extracted data, terminological problems are identified **202**. Certain terms that have been used by several phonological linguists in their attempt to define and describe the nature of various Arabic sounds have proved invalid; i.e. whereas few linguists may include phonemes like /F7/, /R7/, and /X/ into the category of Emphatics, others may include them in the category of pharyngeals. As a result of this non-final consensus, the most appropriate category depending upon their influence on the neighboring vowels was selected. Next, transcription problems associated with the language in question (e.g., Arabic) are identified **204**. In contrast to what the IPA exhibits in using special symbols (ASCII characters) which might cause technical problems if used in the present system; the transcription set was limited to include the characters which can be typed easily on the keyboard. Furthermore, phonological and phonetic units were extracted or collected **206** and a feature set was established based on this information **208**. Next, a representative symbol for the transcription alphabet is selected **210** and a structured source is built **212**. Our structured source consists of

Phonemes, which are divided into three main units: Consonants, Vowels and Semi-Vowels. The unit "*Consonants*" includes a variety of Allophones and Geminations. Allophones may have their own gemination variety. The unit "*Vowels*" has a variety of allophones only, while the unit "*Semi-vowel*" has just gemination variety. The features of these units are determined according to three conditions: Place, manner of articulation in addition to the nature of the sound being voiced or voiceless.

Figure 3 illustrates the composition of the maximal set described in step **110** of Figure 1. Maximal set **300** comprises (but is not limited to): phonemes **302**, allophones **304**, a set of rules governing the selection of allophones **306**, a set of examples **308**, and the transliteration symbols **310**. It should be noted that although the preferred language of this application is Arabic, one skilled in the art could extend the present invention to cover other similar languages. A detailed description of the Arabic phonetic study as per the present invention is given below.

#### ARABIC BASIC PHONETIC STUDY

When starting the research on basic a phonetic study for Arabic language, certain points concerning the nature of the language need to be considered. In other words, the characteristics of Arabic language at different levels (graphemic, morphological, and phonological) need to be considered. To do so, different forms of the Arabic language that can be used as an input for the text to speech (TTS) system need to be identified.

#### ARABIC LANGUAGE VARIETIES

We could distinguish in the Arabic language at least three varieties:

### *Classic Arabic*

- 1) Language of the Holy Qur'an; highly codified since early Islamic period
- 2) Used nowadays only in religious sermons or speeches

### *Modern Standard Arabic (MSA)*

- 1) "Standard": Highly codified language, grammatically identical to classical Arabic, although case ending is not usually pronounced.
- 2) "Modern": Lexically adapted to modern times (e.g., lexicon innovations, loan words).

### *Colloquial Arabic*

- 1) Arabic dialects: Natively learned varieties that are used in informal situations and in the everyday communication of a geographically defined community.

Since the input for the TTS system is text, it is clear that the target language for TTS should be modern standard Arabic (MSA).

### **Arabic Language Has Distinctive Features**

Arabic letters need to be transliterated, in other words, they need to be represented by Roman alphabets in such a way that there is a one-to-one mapping between the two character systems. There is a need to not only transliterate characters, but diacritics also. Therefore, Arabic distinctive phonetic groups were created. For example, as illustrated in Figure 4:

- 1) Pharyngeal phonemes like / t%K/ , / D%K/ , and / d%K/ were created.
- 2) Emphatic phonemes like / F7/ , /R7/ , and /X/ were created.

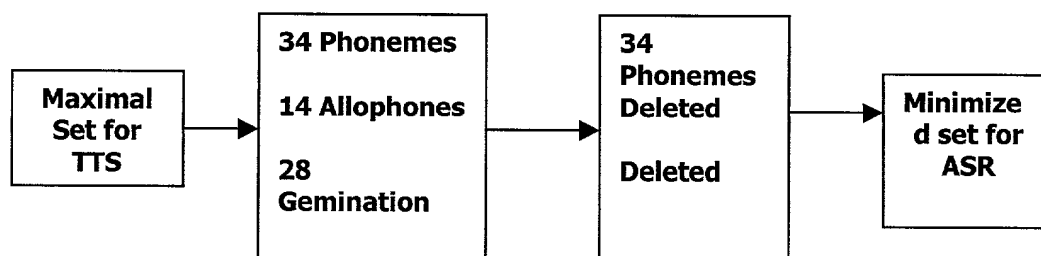
Furthermore, Arabic language has various kinds of allophones:

- 1) The gemination of all consonants.
- 2) Normal allophones, like pharyngealized allophones of certain consonants, and the varieties of vowels.

Finally, Arabic language has a more distinctive syllabification and lexical stress system than any other language. As a result of the basic phonetics associates with the Arabic language, a maximal set is created, that contains all phonemes, the allophones of the language, a preliminary set of rules governing the selection of allophones, a set of examples, and a transliteration symbols.

### **Reduction Of Maximal Set**

The reduction of maximal set for the TTS and the ASR phonetic sets are described in details below:



#### *TTS Phonetic Set*

In the phonetic set for the TTS system, all the phonemes and allophones with which any given text message can be conveyed is found. For example, i) all the allophones for the vowels are identified; ii) allophones that represent any borrowed word in Arabic are identified, and iii) in the case of gemination, add symbols to

represent the phoneme when it is geminated. Thus, geminated phonemes represented by doubling the original symbol, are represented by a new symbol.

### *ASR Set*

For speech recognition, when the system recognizes the three varieties of long *fatha*, all of them are converted to *Alif*, for example, in words like /nE.0Im/, /n1.qId/ and /n2.6Ig/; the vowels E, 1, and 2 are represented by the grapheme Ç. While three varieties for the short *fatha* will be converted into the diacritic *fatha*, for example, in words like /ge.le.se/, /fa.qad/ and /qAws/; the vowels e, a, and A are not represented in orthography. Both the geminated and non-geminated consonant will be represented by the same grapheme.

Only few allophones are added to recognize certain pronunciation varieties due to the country that the speech recognition system is developed for. Thus, in conclusion, the automatic speech recognition (ASR) set is less in number than the text to speech (TTS) set, thereby reducing the memory consumption in the resident computer system enabling easier storage of the compact set of phonetics.

The above enhancements for a compact model to transcribe the Arabic language acoustically based on a well-defined basic phonetic study are implemented in various computing environments. For example, the present invention may be implemented on a conventional computing equipment, a multi-nodal system (e.g. LAN) or networking system (e.g. Internet, WWW, wireless web). All programming and data related thereto are stored in computer memory, static or dynamic, and may be retrieved by the user in any of: conventional computer storage, display (i.e. CRT) and/or hardcopy (i.e. printed)

formats. The programming of the present invention may be implemented by one of skill in the art of automatic speech recognition (ASR).

A system and method has been shown in the above embodiments for the effective implementation of an Expanding Dictation Vocabulary. While various preferred embodiments have been shown and described, it will be understood that there is no intent to limit the invention by such disclosure, but rather, it is intended to cover all modifications and alternate constructions falling within the spirit and scope of the invention, as defined in the appended claims. For example, the present invention should not be limited by software/program, computing environment, or specific computing hardware.